AAS 25-901

# MW-NERF: MULTI-WAVELENGTH NERF MODELS FOR SPACECRAFT MODELING IN SHADOWED ENVIRONMENTS

**Logan Selph*** **and John R. Martin†**

The future of Neural Radiance Fields (NeRFs) within aerospace applications relies heavily on their ability to cope with the hostile observing conditions that permeate space-based operations. While existing NeRF models can handle many real-world scenes with extremely fine detail, they still lack robustness to the extreme lighting conditions present outside of Earth's atmosphere. While optical light may not contain all the information required to faithfully reconstruct a scene, we propose that this problem can be tackled by creating a new NeRF architecture that can process images taken at different wavelengths, where vital shape information may be readily available. This work explores our new proposed model: multi-wavelength NeRF (MW-NeRF), along with a large realistic dataset suite containing simple to complex satellite geometries with multiple different lighting conditions, taken over several different target ranges. Early analysis of our model shows that dark regions in these datasets prove difficult for optical training alone to overcome, and that better underlying shape models can be learned when provided with multi-wavelength datasets.

## INTRODUCTION

As global interest in exploring and operating in space continues to grow, the safety of future missions and the security of national interests must keep pace. Adequate Space Situational Awareness (SSA) is paramount for ensuring the success of missions that operate in these poorly characterized and dangerous environments. SSA relies on tools and infrastructure to track and understand objects operating in orbit. These objects include cooperative and non-cooperative spacecraft, debris, celestial bodies, and others. In all cases, these objects need to be carefully monitored to avoid purposefully or inadvertently jeopardizing current missions or their operations. Spacecraft operators must contend with limited information about other cataloged satellites and debris.[1] While rough state information about these objects is maintained through ground-based observation, other necessary details, such as shape and size, are often unknown. This forces operators to make assumptions which may impact mission resources and safety — e.g. unnecessarily maneuvering to avoid a poorly characterized threat.[1]

While spacecraft operators have many tools at their disposal that can generate useful shape models from images taken by their vehicles, most methods carry serious drawbacks that can make their use on most missions either burdensome or impossible. This mostly has to do with the computational burden and large footprints that typical shape modeling frameworks carry with them. Some missions must be designed to have the time and capacity to send large amounts of data to ground

*NSF Graduate Research Fellow, Aerospace Engineering, University of Maryland, College Park.
†Assistant Professor, Aerospace Engineering, University of Maryland, College Park.

stations, which then spend days to months generating the final model, for this very reason. Unfortunately, not every mission has such resources: small satellites with data downlink constraints or missions that require quick model outputs demand lightweight and on-board solutions.[2] Our research focuses on developing fast and lightweight machine learning models that can reconstruct 3D scenes containing poorly characterized objects to enhance SSA.

Two methods are frequently used for 3D shape modeling: reconstruction[3] and template matching.[4] The former relies on using a set of 2D images of the target to reconstruct either the entire scene, including the target, or just the target itself, with implementations including analytic, numerical, and machine learning methods.[5,6] Reconstruction in space is particularly challenging because imaging conditions are often far from ideal.[6] This means that reconstruction methods must be robust to extreme lighting conditions, reflections from non-Lambertian materials, motion blur, and other prevalent issues within space-based observation. While reconstruction methods make no prior assumptions about the target's geometry, template matching relies on precise information regarding the object's shape and structure to apply feature mapping.[7,8] Alternatively, machine learning implementations have proven to be powerful solutions that can quickly and accurately estimate the 3D shape models of these objects within their local environments. Among these tools, the recently established Neural Radiance Field (NeRF) stands out.

NeRFs employ a multi-layer perceptron to represent a scene as a field of density and view-dependent radiance values.[9] Specifically, NeRFs construct a mapping from position and viewing direction in space to density and color values. To generate a novel viewpoint—that is, a new image rendered from a perspective not available in the original dataset—each pixel value is estimated by rendering the relevant points in the learned field. The result is a continuous, learned field of density and color values that can be used to render images of the object from any arbitrary viewing geometry. Because this method learns the scene without requiring any prior knowledge or assumptions, it has proven to be an exceptionally powerful tool used in a variety of applications.[10]

This work aims to extend NeRF models to create a fast and accurate method for rendering scenes with spacecraft in difficult lighting environments. Specifically, we investigate the issue of shadows impeding a NeRF's ability to render a scene effectively. As a solution, we propose MW-NeRF, a framework that uses NeRFs trained on data obtained at multiple wavelengths to better inform training in the visual band—an approach that, to our knowledge, has not yet been explored. Our goal is to develop a reconstruction method that is robust to adverse lighting conditions and capable of reliably learning scenes from images captured by spacecraft at realistic working distances.

## BACKGROUND AND RELATED WORK

### Scene Modeling in Astrodynamics

Traditional shape reconstruction methods for objects in orbit include structure from motion (SfM) and point cloud-based techniques.[11] SfM algorithms detect and match features between 2D images to accurately recover camera poses, then leverage the resulting parallax between viewpoints to triangulate 3D points and produce precise depth reconstructions of the scene. In space-based environments, SfM reconstructions commonly suffer from the presence of holes in the final renderings, especially on low-feature or non-Lambertian surfaces where the feature mapping can struggle. As a result, 3D modeling of objects in orbit has largely continued to rely on stereophotoclinometry (SPC) based models.

SPC combines stereo parallax with photoclinometry to transform 2D images of a scene into a

3D digital terrain map (DTM).[12] This method can generate multiple DTMs to effectively map an entire object, and has been used to model dozens of celestial bodies in our solar system, including the Moon, Ceres, Bennu, and Itokawa.[13] SPC is excellent at reconstructing 3D models of celestial bodies, but it does rely on very large datasets and takes weeks to months of computation time to produce its best results.[13]

Other approaches to retrieving 3D scene information for Earth-orbiting satellites include using ground-based inverse synthetic aperture radar (ISAR) observations. This can also be combined with bidirectional analytic ray tracing (BART) to further assist the extraction of 3D information from radar images. Several studies have successfully applied ISAR and BART both to generate 3D scene information and to estimate satellite attitudes.[4,5] These methods are capable of learning sparse shape models from ground based observations, but cannot generate continuous representations of scenes with fine detail. More recently, pre-trained convolutional neural networks (CNNs)—architectures originally developed for image classification and detection—have been adapted to predict full 3D shapes from single optical images. The approach in reference 7 trains a CNN on a library of spacecraft geometries and then infers complete 3D models given just one view. While this delivers impressive generalization across different vehicle classes, its performance hinges on the diversity and quality of the training dataset, and it may struggle when presented with unfamiliar geometries.

**Neural Radiance Fields**

NeRFs are a class of machine learning models that learn a continuous 3D model of a scene from a finite set of 2D images. Once the model is trained, it can synthesize new images from any arbitrary viewpoint while accurately reproducing the scene's lighting and preserving its geometric structure. The models are trained by sampling points along rays spawned from each training image, and integrating the corresponding density and radiance values at each field point along the specific ray to generate a synthetic pixel value. The model is penalized for inaccurate rendering of the original 2D images, and its weights are updated using stochastic gradient descent until the model converges on a set of field values that can reliably reproduce the training images. Using the same ray-marching strategies, users can then query the learned neural radiance field to generate new images from novel poses not previously seen during training. The result amounts to a photorealistic, continuous approximation of the entire scene.

Explicitly, this training process would involve numerically or analytically computing the integral that produces the final color value for a ray via:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt, \text{ where } T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds\right). \tag{1}$$

In Equation (1) $t_n$ and $t_f$ measure the starting and stopping points along the ray, $\mathbf{r}$ is the sampled position in 3D space, $\mathbf{d}$ is the normalized viewing direction, $\sigma$ is the position dependent density, and $\mathbf{c}$ is the position and view dependent color. $T$ is the accumulated transmittance along a ray, representing the probability that the ray travels some distance without hitting another particle.

Calculating this integral in its entirety would be computationally burdensome, so it is approximated by sampling points along a ray where light scatters. Instead of computing the full continuous field, a simpler mapping is learned by approximating the integral through a finite number of samples along the ray:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i(1 - \exp(-\sigma_i \delta_i))\mathbf{c}_i, \text{ where } T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j \delta_j). \tag{2}$$

In Equation (2), $\delta$ represents the distance between the current sample and the next.

Optimization is achieved by comparing the model's performance on a per-ray basis, rather than per sample, since the ground truth of the entire radiance field is not known. To do this, the final renderings from the above equation are compared to the ground truth pixel values, resulting in the loss function:

$$\mathcal{L} = \sum_{\mathbf{r} \exists \mathcal{R}} ||\hat{C}_\theta(\mathbf{r}) - \hat{C}(\mathbf{r})||_2^2, \tag{3}$$

where $\hat{C}_\theta$ is the output ray value of the network, and $\hat{C}$ is the ground truth value.

Since the introduction of the original NeRF model in 2020, dozens of new methods have emerged to address various drawbacks and expand its application to different fields. Of these advancements, some of the most important have to do with training speed,[14,15] accuracy in real-world scenes,[16–18] estimating camera poses,[19] very large scenes,[20,21] unbounded scenes,[22,23] and time-varying scenes.[24,25] In the context of this paper, the most important of these additions include optimized CUDA kernels for speed, appearance embedding for real-world scenes, and proposal sampling for unbounded scenes.

*CUDA Kernels* The original NeRF model was extremely slow to train and query for a few different reasons, the most notable being its lack of optimal GPU utilization. Instant-NGP introduced optimized CUDA kernels in 2022, transforming train times that previously took up to a day into mere minutes or seconds.[15] This is achieved by squeezing the entire NeRF training process (including MLP layers, grid interpolations, stored feature vectors, etc.) into a single CUDA kernel. By storing everything on video memory and avoiding large GPU launch counts for repetitive tasks, massive amounts of computation overhead are bypassed entirely. The final result allows us to train NeRFs hundreds of times faster compared to the original models produced in 2020.

*Appearance Embedding* Real-world data rarely has ideal, static lighting conditions. Since the original NeRF model was designed more as a proof-of-concept using datasets made in Blender, it was naturally ill-equipped to tackle datasets with changing lighting, such as with images taken during different parts of the day, or clouds intermittently casting shade in different areas. This is because the fully-fused MLP in the original NeRF model learns entire scenes from explicit five-dimensional inputs (a 3-D position and a 2-D pointing direction). The original color field optimizes a view-dependent color representation for the final learned scene. While this is well-equipped to learn features that vary solely in the direction that you observe them, such as reflective surfaces, it struggles to handle unique lighting profiles that cannot be described by viewing direction alone. Real-world datasets are likely to have such full scene lighting differences between images, making it much more difficult for the original model to reconcile these conflicting viewpoints, and overall damaging the final learned reconstruction. W-NeRF solved this problem by adding a learned appearance embedding vector to each training image, giving the color field additional image-dependent parameters to tune that allow it to represent continuous representations for inter-image lighting differences.[16] This results in a model that can accurately learn real-world reconstructions from datasets that depict the same scene in many different contexts.

*Proposal Sampling* The process of approximating a ray requires many samples to be taken along its path until the surface of the intersected object has been fully learned. Without any guidance this process can take quite some time to converge on an accurate render, since random sampling will mostly yield points in empty space or points inside the object. This is not only burdensome for bounded scenes focusing on single objects, but this quickly becomes intractable for large, unbounded scenes with objects scattered across all scales. Mip-NeRF 360 introduced the proposal sampler, which simultaneously optimizes a lighter version of the same NeRF, except it solely learns the underlying density field.[22] This light field takes in a uniform sample of points along a ray and outputs the density weights associated with each of those points. These points then make up a density histogram, where new candidate sample points can be chosen through inverse-transform sampling. The new points, which are concentrated towards objects in the scene rather than empty space, are then used in the forward pass for training the base NeRF model. This proposal network is then optimized using the KL divergence between its density field outputs and those of the base NeRF model, meaning that the proposal network is optimized based on where the NeRF network is placing geometry in the scene. This drastically improves sample efficiency, which greatly speeds up training by feeding the model more information that directly contributes to learning the underlying geometry of the scene.

Since 2020, the state of the art for NeRFs has advanced dramatically—from requiring several hours of training on simple, bounded datasets to training in seconds and rendering entire unbounded scenes with exquisite detail. With these improvements in both speed and accuracy, NeRFs are now being tested in various fields with great success. In astrodynamics, for example, recent work has begun to explore the utility of NeRFs for guidance, navigation, and control, particularly in applications focused on the 3D modeling of spacecraft and celestial bodies.

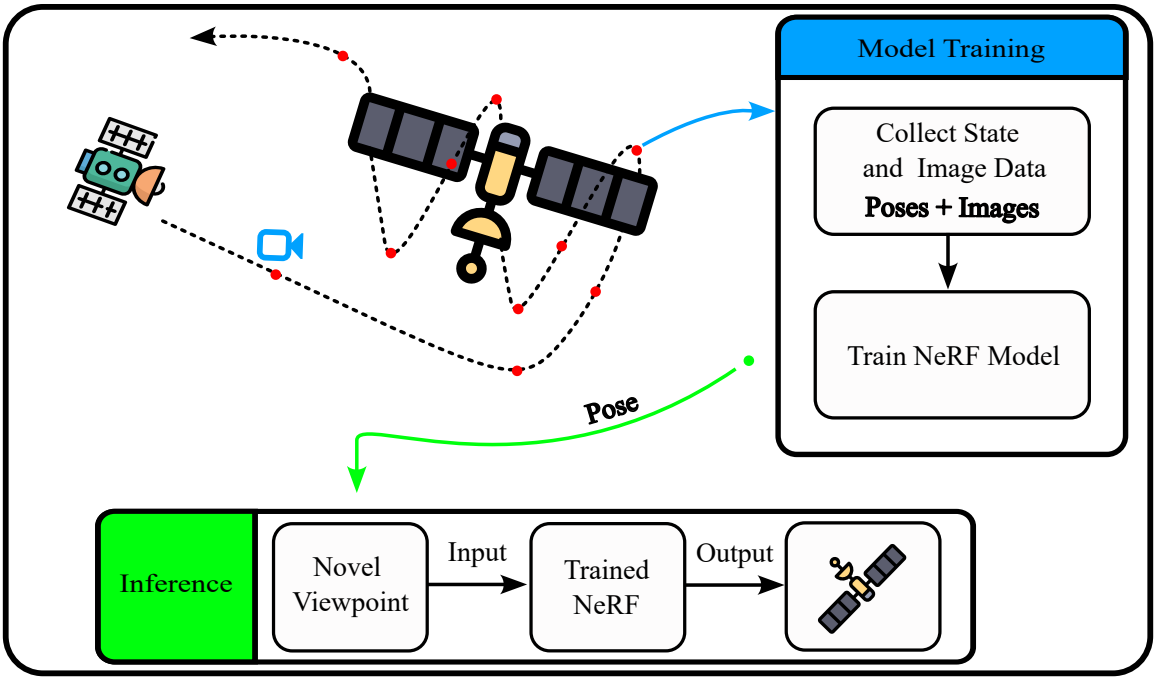



**Figure 1. The baseline procedure for collecting data for training and rendering novel views from a NeRF model. Poses represent the position and normalized viewing direction of the cameras for each observation or novel viewpoint.**

### NeRFs in Astrodynamics

Past efforts have explored the use of NeRFs and other neural network based architectures to generate 3D models of spacecraft and celestial bodies from sparse sets of images. These efforts include using NeRFs to model asteroids—demonstrating that they can be faster, more accurate, and require less data than current state-of-the-art techniques like SPC.[26] Other studies have concentrated on reconstructing spacecraft from images taken on-orbit, exploring limitations such as the quantity of images, complex lighting conditions, and non-ideal imaging effects like motion blur.[6,27,28] These investigations highlight the potential of NeRFs for generating detailed 3D models of various objects found in-orbit. Still, challenges persist in maintaining NeRF quality under non-ideal observation conditions. Many NeRF models suffer when trained on data captured in spacecraft environments, where images are taken from long distances, lighting is highly variable, and targets are in rapid motion. Future work must address these challenges to fully exploit the potential of NeRFs in realistic mission scenarios.

In addition to pure scene reconstruction, recent studies have explored the use of NeRFs for proximity operations and target pose estimation. Reference 29 employs two separate NeRF models: one (W-NeRF) to build a model of the real-world scene using pre-processed images of a target body, and another (iNeRF) to extract the spacecraft's state by optimizing its pose in the learned scene. As new images are captured, iNeRF "inverts" the process—optimizing the camera position within the W-NeRF-generated scene to best reproduce the new image. Other research has focused on using NeRFs to generate detailed 3D models of target spacecraft, which can then be integrated into existing pose estimation software.[30] For instance, a study using a newer K-planes model demonstrated promising results in estimating both the scene and the target spacecraft's pose.[25] Despite these successes, challenges remain: hostile lighting conditions continue to negatively impact NeRF performance, and the optimal dataset for reconstructing a target scene is typically one that is captured at close range.

### Thermal Imaging

Adverse imaging conditions, such as intense shadows produced by sunlight, present significant challenges across many state estimation and model reconstruction techniques. Recent research has explored using alternative wavelength bands—particularly the infrared band—for visual navigation when the visual band is compromised. Although infrared images generally have lower resolution, they consistently provide reliable and useful information that complements visual data.[31] While thermal imaging may struggle with fine edge detection, it is decidedly preferable to having no information at all.

There has been some exploration into using NeRFs with inputs from different wavelength bands, but no work to date has specifically coupled these bands to provide shape priors when the optical wavelength is insufficient. The ThermalNeRF model was among the first to effectively combine RGB and thermal imaging into a single NeRF scene.[32] This model expands upon the base Nerfacto model provided by nerfstudio, allowing it to learn RGB and thermal fields simultaneously by optimizing two separate models for each color and density field. To maintain consistency between the two scenes, ThermalNeRF connects the separated fields through additional regularizers specified in their loss function. This model learns accurate representations for both the underlying RGB and thermal fields present in the training dataset, allowing it to represent objects that may have been entirely obscured from view in either ground-truth dataset. While the ThermalNeRF publication does not explore challenging lighting environments, they do demonstrate that RGB-T datasets can be

effectively combined to reveal information that would otherwise be inaccessible when using RGB or thermal alone.

Another closely related work involves a single NeRF (MultiBARF) that combines inputs from two separate wavelengths to produce a model yielding both visual and thermal outputs.[33] MultiBARF is built from the original Bundle-Adjusting Neural Radiance Field (BARF) model, which tackles the issue of refining camera pose estimation in the same way as modern simultaneous localization and mapping (SLAM) algorithms—using bundle adjustment.[34] MultiBARF demonstrates that a single NeRF model can simultaneously learn scene representations from two sensors operating at separate wavelengths. This model employs a single density field, which keeps the underlying geometry of the scene consistent across color channels. MultiBARF's results found that when trained synchronously, depth estimates improved compared to thermal-only training. When trained asynchronously, one sensor dataset after the other, there was degradation of the final field in each case. Overall, MultiBARF implemented a NeRF framework capable of learning datasets containing images in different wavelengths, without the need for extra camera calibration or precise initial pose estimates. No study has taken this work further to explore how thermal data can inform the geometry of what optical data fails to capture, and how this can be leveraged to enhance 3D model reconstruction for objects observed in space. Further research is needed to optimize thermal scene representations and to leverage them as shape priors that enhance the performance of visual band NeRFs in challenging lighting environments.

## METHODS

We seek to develop candidate mechanisms to assist in reliably training NeRF models under challenging lighting conditions, such as those found in orbit. Figure 1 illustrates the traditional end-to-end procedure for training a NeRF model. In this baseline method, a set of images capturing different parts of a scene is used to learn a continuous volumetric representation. However, in space-based observation, significant portions of the scene can be missing or poorly illuminated due to the presence of shadows. With the sun often serving as the primary (or sole) light source, many regions of the target object may remain entirely unlit. In such cases, a NeRF model is challenged in two key ways: incomplete learning and model divergence. In the case of shadows, the learned geometry in these areas suffers greatly as the NeRF has no information to refine its estimate. This lack of data results in poor performance in shadowed areas, having the potential to impact training in visible areas as well, causing the entire model to diverge in the worst cases. In order to be robust to these challenging viewing conditions, the original NeRF framework must be expanded to accept additional information that reinforces its learning of a scene.

To overcome these limitations, we propose MW-NeRF, an approach that leverages multi-wavelength imaging to improve the reconstruction of spacecraft under adverse lighting conditions. The core idea is to train a single NeRF model that approximates color in two separate wavelengths, similar to MultiBARF and ThermalNeRF, but based on a K-Planes model architecture. This model will lean on visible spectrum data for well-lit portions of the scene, switching to data from a separate wavelength when it would serve to improve the final reconstruction accuracy. This provides the best of both worlds for celestial body and spacecraft modeling—describing the visible portion in high detail and constraining the remaining shadowed areas using data from a different wavelength.

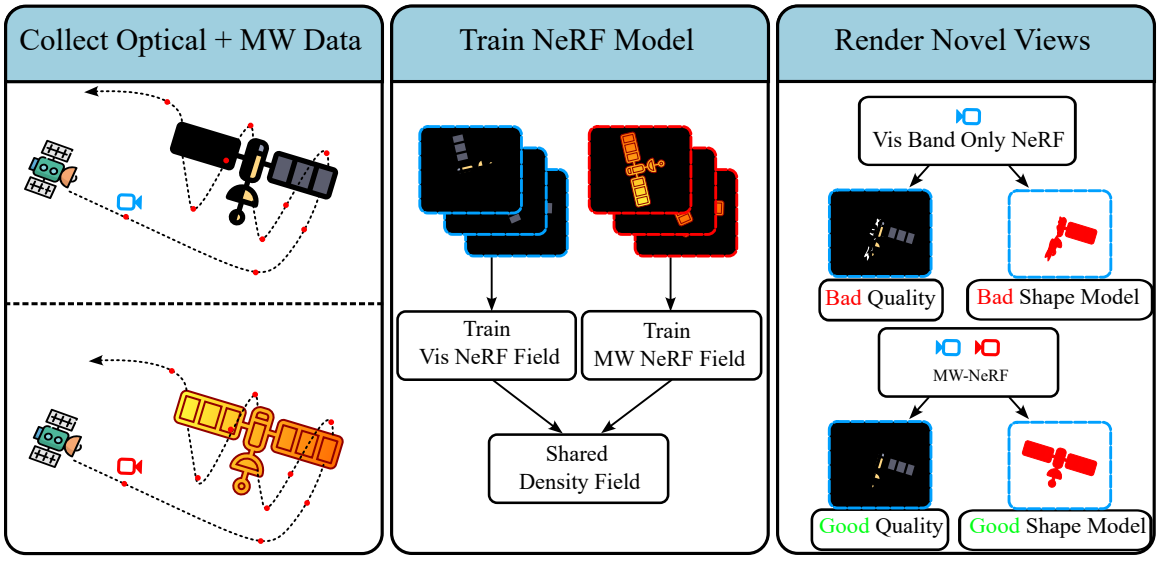



**Figure 2. The proposed modified procedure for MW-NeRF. This approach involves training a single NeRF model using data from different wavelength bands. The part of the model optimized on an alternative wavelength, where desired features are more visible, provides a learned shape prior that informs the visual band in shadowed or low-information areas.**

### K-Planes

The K-Planes NeRF model architecture is fast, compact, and can handle changing lighting conditions. This model is also capable of reconstructing both static and dynamic scenes, meaning it can render objects moving in a scene over time. K-Planes achieves this by its clever choice of feature encoding, opting to utilize "d-choose-2" planes to represent d-dimensional scenes. In the static case, this results in a "tri-plane" representation with the scene being represented using the xy, yz, and zx coordinate planes. Each of these planes is broken down into a pre-defined spatial resolution where learned feature vectors can be stored and interpolated when querying any point on these grids. Additional features from W-NeRF and Mip-NeRF-360 were also included in this architecture, including appearance embedding and proposal network sampling. K-Planes does boast the ability to use linear decoders rather than MLPs to learn accurate scene representations, but its best results use MLPs, which also benefit from the CUDA kernels introduced in Instant-NGP. The loss function for this is computed from the same color rendering algorithm as the original NeRF, and promotes feature smoothness in planes and spatial dimensions through a spatial total variation regularization term $\mathcal{L}_{TV}$:

$$\mathcal{L}_{TV}(\mathbf{P}) = \frac{1}{|C|n^2} \sum_{c,i,j} \left( \|\mathbf{P}_c^{i,j} - \mathbf{P}_c^{i-1,j}\|_2^2 + \|\mathbf{P}_c^{i,j} - \mathbf{P}_c^{i,j-1}\|_2^2 \right), \tag{4}$$

where $\mathbf{P}_c$ are the $c$ different planes, $C$ is the set of all 2D feature planes, $n$ is the resolution width of the planes, and $i, j$ are the indices of the different grid points given by the resolution of the plane.

**MW-NeRF**

MW-NeRF is built on the original K-Planes model, tailored to reconstruct space-based scenes using single-channel 32-bit float datasets. Our NeRF models are all built using nerfstudio, an open-source, developer-friendly, and highly collaborative repository. The current build of MW-NeRF has a few changes that allow it to work well with space-based datasets besides being able to use single-channel, high-range images — they are as follows.

*Pixel Sampling* While NeRFs do require objects to be resolved to learn a scene from scratch, it is still important to recognize that observations may not always occur at favorable distances. Although objects can still be resolved at long distances, they may take up a tiny portion of the final image, causing the randomly sampled rays the NeRF attempts to learn the scene from to be dominated by background pixels rather than the object of interest. In our formulation, we assume that there is some rough estimate available to satellite operators indicating their distance to the object being observed. These distances allow us to implement a well-informed axis-aligned bounding box, or scene box, which encases the area of interest being reconstructed. We have added a new pixel sampling method that uses the width of our scene box to mask out regions of images with unnecessary background information. Fine-tuning these scene box parameters allows the MW-NeRF model to concentrate its pixel samples on the object of interest, while still sampling some background pixels, which we have empirically found to greatly benefit reconstruction stability.

*Ray Sampling* After sampling rays that intersect the scene box, we also have to guide where along those rays we are interested in pulling initial samples. This comes in the form of calculating near and far planes, which specify the closest and furthest points along the ray that the model is allowed to sample from. We have implemented a minimal ray-tracing algorithm, known as the Slab method, to calculate where each ray enters and exits the defined scene box. First, let the box be defined by its lower and upper corners

$$\mathbf{l} = (l_0, l_1, l_2), \quad \mathbf{h} = (h_0, h_1, h_2)\,, \tag{5}$$

and a ray be parameterized by

$$\mathbf{p}(t) = \mathbf{o} + t\,\mathbf{r}, \tag{6}$$

where $\mathbf{o} = (o_0, o_1, o_2)$ is the origin and $\mathbf{r} = (r_0, r_1, r_2)$ is its pointing direction.

For each axis, the ray intersects the two planes orthogonal to that axis at

$$t_i^{\text{low}} = \frac{l_i - o_i}{r_i}, \quad t_i^{\text{high}} = \frac{h_i - o_i}{r_i}. \tag{7}$$

We then form the "near" and "far" extents within that slab as

$$t_i^{\text{near}} = \min\big(t_i^{\text{low}},\, t_i^{\text{high}}\big), \quad t_i^{\text{far}} = \max\big(t_i^{\text{low}},\, t_i^{\text{high}}\big). \tag{8}$$

The overall entry and exit parameters for the box are then

$$t^{\text{near}} = \max_i\, t_i^{\text{near}}, \quad t^{\text{far}} = \min_i\, t_i^{\text{far}}, \tag{9}$$

Providing the $[\,t^{\text{near}},\, t^{\text{far}}]$ that bounds our initial uniform samples that are fed into the first pass of the proposal network.

*Multi-Wavelength Channels* Currently, MW-NeRF supports datasets that have up to two different channels. To accommodate the additional channel, we have added an additional color network dedicated to learning the color output for that channel. The underlying density field is shared between the two networks, promoting them to learn a shape model that conforms to the data captured in both channels, without the output color values interfering with each other. The latest version of MW-NeRF allows users to tweak how much each channel is weighted in the loss function, should one channel express desired features more than the other.

The key innovation of MW-NeRF lies in the integration of these two channel modalities. The multi-wavelength portion, which benefits from enhanced feature visibility, is used to generate a shape prior that informs and guides the visible band NeRF during training. This dual-mode strategy offers several advantages, the largest being increased robustness to the challenging lighting conditions present in space environments. By incorporating observations from multiple sources, the risk of becoming limited by data sparsity diminishes greatly, increasing the consistency and reliability of the model as a whole. Informing these previously problematic areas also significantly decreases the likelihood of model divergence, even in extremely dark scenes. A comprehensive, multi-wavelength model can also serve as a foundation for advanced future applications such as dynamic relighting, where understanding the complete shape and structure of the spacecraft is crucial.

Figure 2 summarizes the modified procedure for training MW-NeRF. The diagram shows how the multi-wavelength and visible bands are trained in parallel, with the learned shape information from the MW branch being used to refine the visible branch. Overall, the proposed MW-NeRF framework is designed to leverage complementary spectral information to overcome the inherent limitations of single-band imaging in space. By doing so, it promises to deliver a more robust and accurate reconstruction of spacecraft, even under the challenging illumination conditions found in orbit.

**MW-NeRF Dataset**

While MW-NeRF will be designed to accept images across many wavelengths, our first exploration involves the complementary pairing of the visible and infrared wavelengths. We will use synthetic datasets of two canonical spacecraft geometries—a custom CubeSat ("Boxsat") and a Hubble telescope model—running a photorealistic rendering engine to simulate both visible and infrared imagery under controlled, physically accurate conditions. For each model, the rendering engine will compute scene radiance by tracing solar illumination and thermal emission across user-defined lighting geometries (varying solar incidence and phase angles), platform-target distances (from kilometers to meters away from the target), and spectral channels spanning visible and infrared wavelengths. These multi-wavelength image pairs will then train our MW-NeRF: the IR branch provides a robust shape prior in deep shadows, while the visible branch captures high-fidelity surface textures in well-lit regions. By systematically varying angle, range, and wavelength, we will assess MW-NeRF's ability to fuse disparate spectral data, quantify reconstruction accuracy in both bands, and demonstrate enhanced robustness in the extreme lighting environments typical of celestial body and spacecraft imaging. The following section, covering our current results, will quantify the performance of the MW-NeRF model architecture first when training on datasets taken in only the visible spectrum, followed by dual-channel training.

In our initial experiments, we focus exclusively on optical data for the two canonical spacecraft geometries. For each geometry, we generate images using a uniform 614-point spherical sampling grid under identical solar illumination. Five separate datasets are produced per geometry, each

corresponding to one of five discrete viewing distances. Because the appropriate distance depends on target size, the Boxsat model ranges from 20m to 300m and the Hubble model from 50m to 1km. Figure 3 shows a single example for each of these spacecraft across their range of viewing distances. Our exploration of these datasets will focus on a single trained model per spacecraft geometry, each trained on the full dataset of 3070 images.

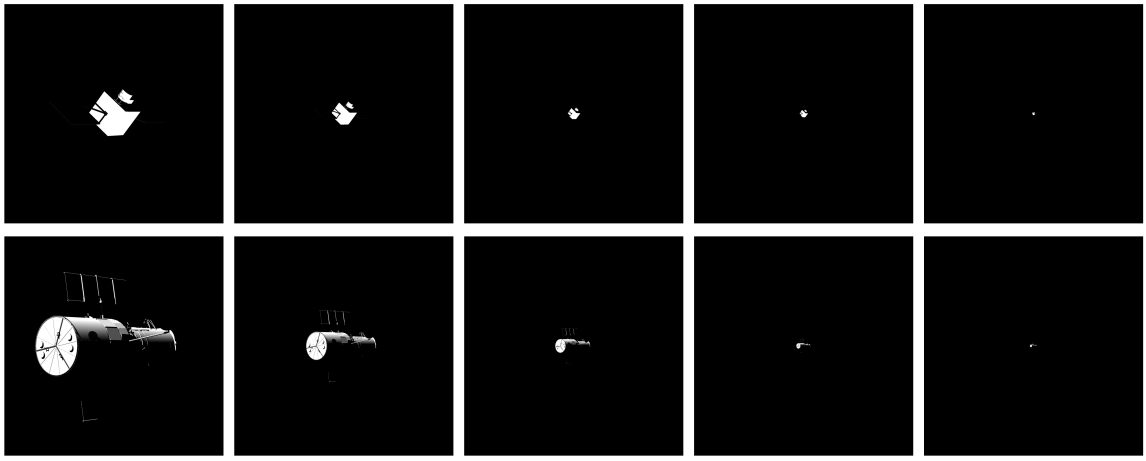


**Figure 3. A single viewing direction from the two canonical spacecraft geometries, spread across their five different viewing distances.**

## PRELIMINARY RESULTS

Our current results first focus on exploring the baseline performance of the MW-NeRF architecture when training on just a single channel. In this exploration, we will show the view and distance-dependent performance for each of the canonical spacecraft geometries defined previously. More specifically, we will point out shared points of difficulty across all spacecraft geometries and places where single-channel models fall short. We will then move on to the current results being produced by the two-channel MW-NeRF model, outlining its current state, what changes need to be made, and what we expect to improve once it reaches its final version.

### Performance Metrics and Figures

Since NeRFs reconstruct a continuous 3D scene and render novel 2D views, their quality is commonly assessed with image–comparison metrics from computer graphics. We use the two most popular: the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM).

PSNR is a log-scaled transform of the mean squared error (MSE) between a reference image $I$ and a reconstruction $\hat{I}$:

$$\mathrm{PSNR} = 20 \log_{10}\left(\frac{L}{\sqrt{\mathrm{MSE}}}\right), \tag{10}$$

where

$$\mathrm{MSE} = \frac{1}{N}\sum_{p=1}^{N}(I_p - \hat{I}_p)^2, \tag{11}$$

$N$ is the number of valid pixels, and $L$ is the peak possible pixel value. For 8-bit imagery this peak pixel value is capped at $L = 255$. In the case of MW-NeRF, which uses a 32-bit floating point

format, we define this value using the range of pixel values present in the training dataset, choosing $L$ to be the $99.5^{\text{th}}$ quantile of illuminated pixel values in the training set. This choice spans nearly all valid signal without artificially inflating PSNR. Good PSNR values vary based on the type and range of data being represented, with larger values indicating smaller errors between comparison images.

While PSNR is excellent at quantifying overall error magnitude, it is not necessarily indicative of perceptual or structural similarity. SSIM was designed to better correlate with human perception by comparing local luminance, contrast, and structure. For a (typically Gaussian-weighted) window, SSIM is defined as

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{12}$$

where $\mu_x, \mu_y$ are local means, $\sigma_x^2, \sigma_y^2$ local variances, $\sigma_{xy}$ is the local covariance, and $C_1 = (k_1 L)^2$, $C_2 = (k_2 L)^2$ are small stabilization constants (commonly $k_1 = 0.01$, $k_2 = 0.03$). As with PSNR, we use the same $L$ (our quantile-based dynamic range). SSIM ranges from 0–1, with 1 indicating perfect similarity. Because it normalizes by local statistics, SSIM better captures structural fidelity than PSNR; however, it still assumes pixel-wise alignment, can be affected by large global luminance shifts, and depends on the chosen window size.

PSNR and SSIM satisfy our needs for overall error magnitude and structural similarity, but they can be unintuitive to readers unfamiliar with these metrics. To provide a more directly interpretable heuristic, we introduce a third metric: the total intensity percent error (TIPE) between two images. Let

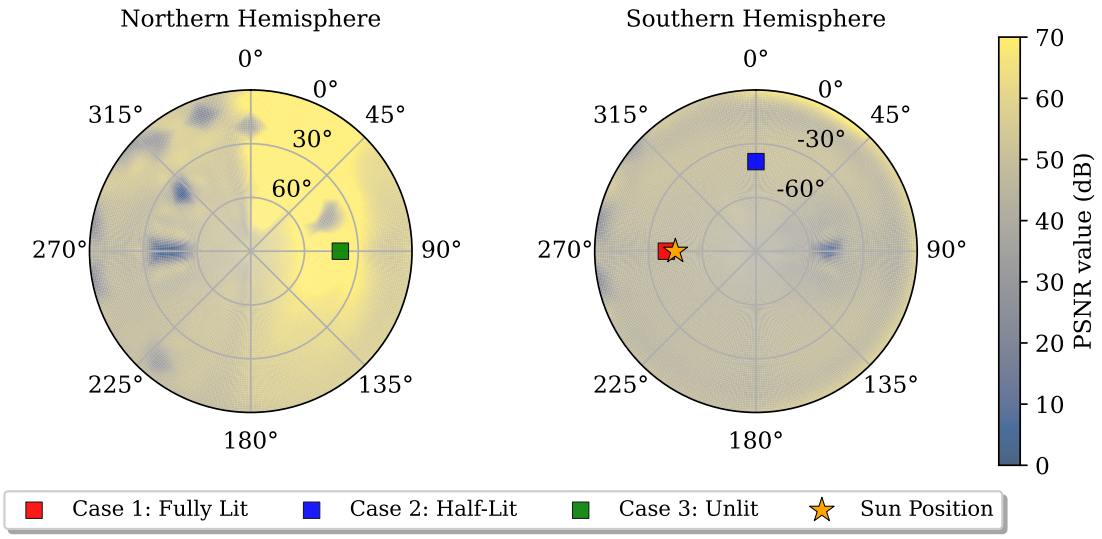
$$\mathcal{I}(I) = \sum_p I_p, \tag{13}$$

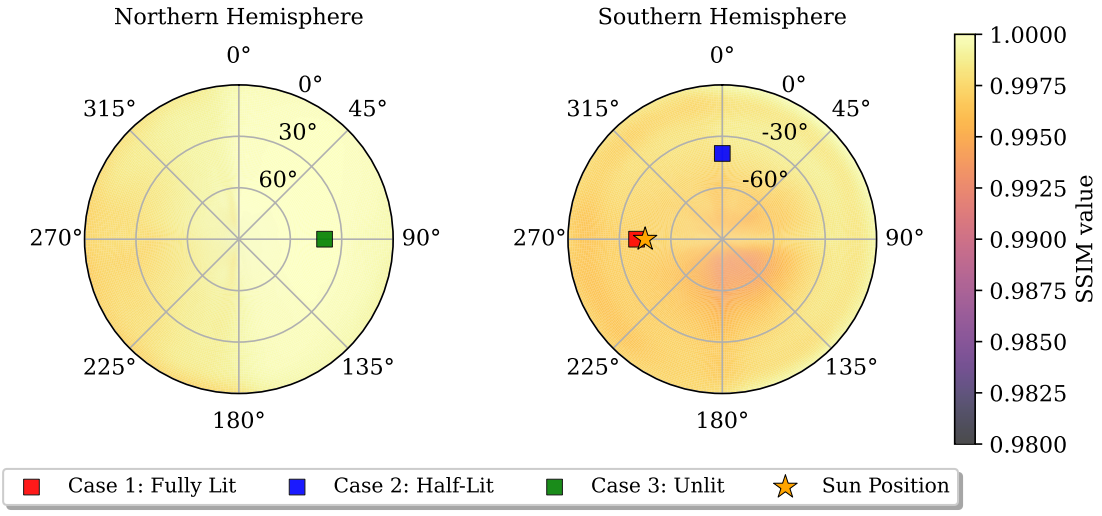
be the total intensity of image $I$ over the set of all pixels in that image. The percent error between a reconstruction $\hat{I}$ and the reference $I$ is

$$\mathrm{TIPE} = 100 \times \frac{\left|\mathcal{I}(\hat{I}) - \mathcal{I}(I)\right|}{\mathcal{I}(I)} \ \%. \tag{14}$$

This scalar value provides a plain percentage describing how much total radiance/energy the reconstruction gained or lost relative to the ground truth. While it ignores the spatial distribution of the error, it complements PSNR and SSIM by offering an easily understood summary of photometric fidelity.

Each metric was computed by comparing the MW-NeRF reconstruction to the ground-truth image at every node of the spherical grid. Figure 4 visualizes these results as polar maps for the northern and southern hemispheres, and outlines the conventions which will be followed by the simplified versions of these figures in the remainder of this paper. Concentric circles indicate elevation, while radial spokes denote azimuth. After evaluating the metric at each grid point, we produced a smooth field using cubic interpolation. Viewpoints from models trained on all available data are highlighted with markers; the specific rendered examples included in the paper are also indicated for reference. All polar plots will follow the same conventions as Figure 4, using the same layouts, color schemes, and color value scales per metric—maintaining consistency for comparisons between different models.



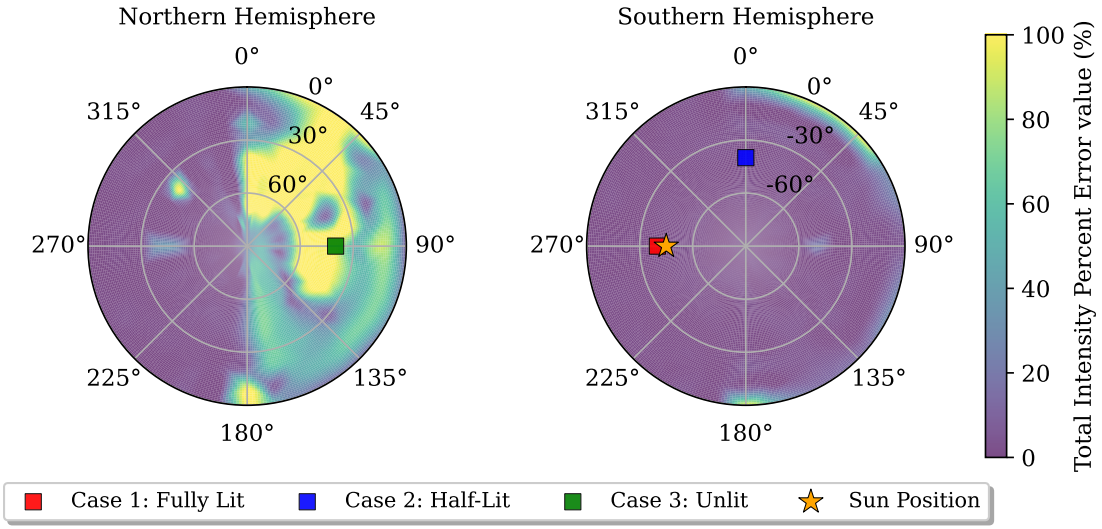




**Figure 4. Interpolated PSNR, SSIM, and TIPE sweeps comparing the ground-truth Boxsat to the visible band MW-NeRF outputs at a distance of 20m.**

### Full Range Training

To begin the exploration of models trained on all 3070 images per spacecraft geometry, we will cover the Boxsat geometry illuminated from a single sun angle at a range of 20m. Figure 4 introduces the polar plots used for each model analyzed in this paper, starting with the close-range results from our fully trained Boxsat model. While SSIM and PSNR show overall high similarity and low error for most parts of the learned radiance field, the TIPE shows how this model struggles in shadowed regions. Because this training dataset contains many shadowed faces, our single-channel MW-NeRF model cannot capture the satellite's full shape.

**Table 1. Boxsat 20m Hemisphere Summary Statistics**

| Metric | Hemisphere | Median | Min | Max |
|---|---|---|---|---|
| TIPE | North | 10.9227 | 0.0081 | 792.503 |
| | South | 1.1978 | 0.0075 | 20.340 |
| PSNR | North | 54.7526 | 1.9329 | 94.4114 |
| | South | 45.0270 | 5.2017 | 49.9770 |
| SSIM | North | 0.999319 | 0.9968 | 0.999997 |
| | South | 0.996956 | 0.9936 | 0.998749 |

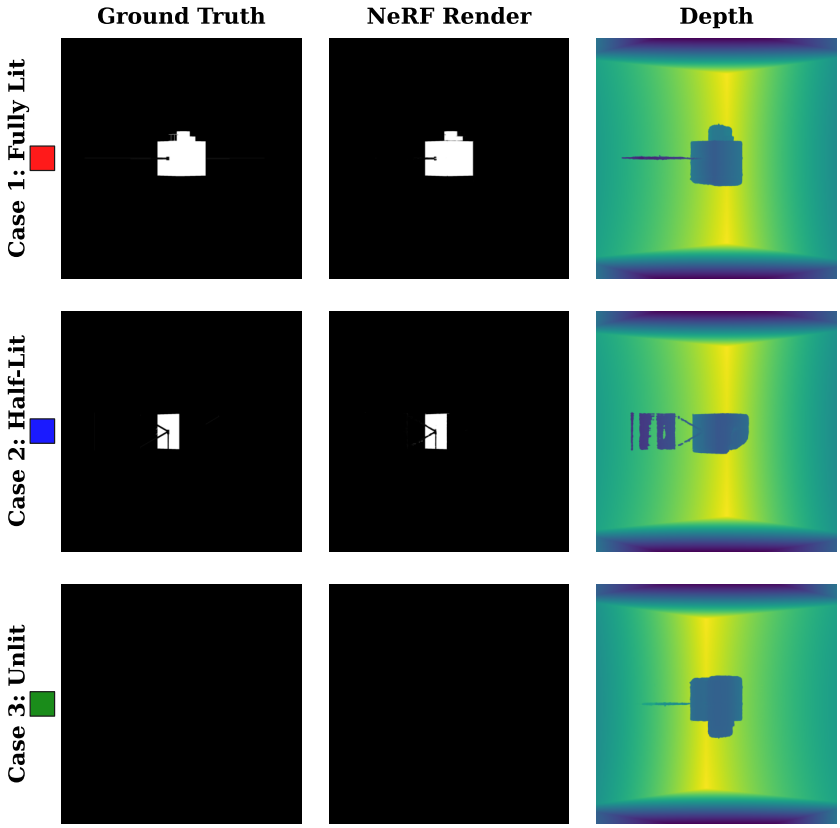
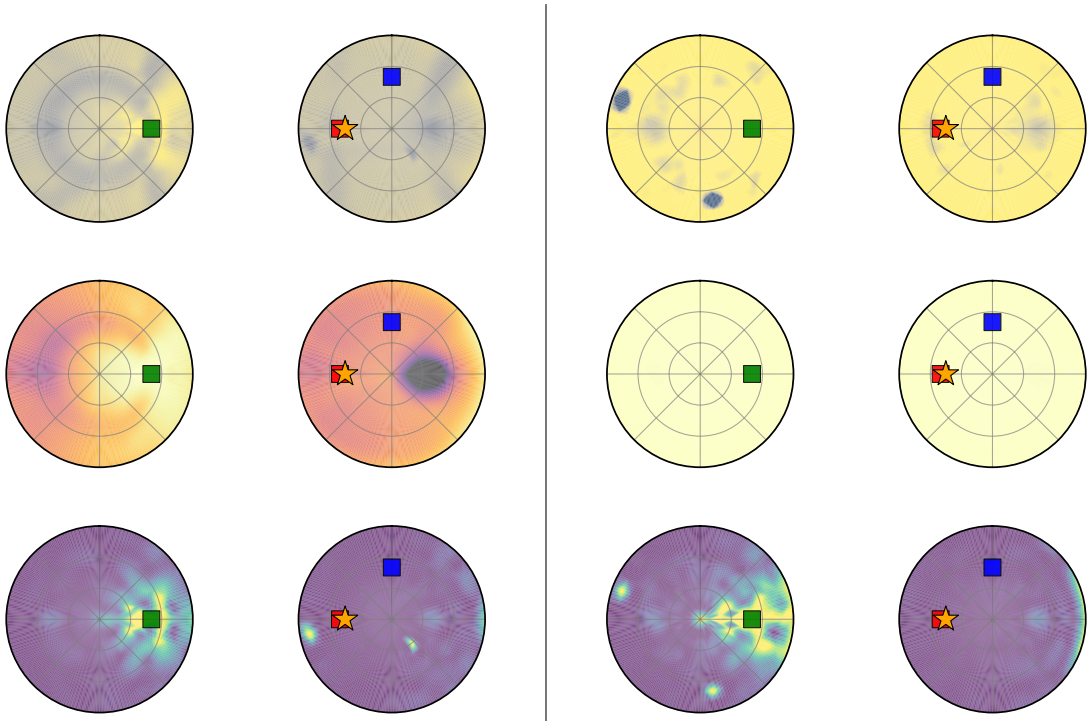



**Figure 5. Side-by-side comparisons of the learned Boxsat across the three lighting conditions labeled in the polar metric plots.**

Figure 5 demonstrates that the visible-band renderings generated from this model are capable

of maintaining a very high degree of similarity compared to their ground truth counterparts, while the depth maps reveal significant portions of the spacecraft that are either poorly learned or not learned at all. This includes some of the smaller illuminated facets jutting out from the sides of the spacecraft, and the very poor reconstruction for one of the solar panels (with the other solar panel being absent entirely). Table 1 outlines the statistics for each of the metrics outlined in the polar plots. Overall, we see that each metric performs well, except for TIPE in the heavily shadowed northern hemisphere. These large max percent error values reach into the hundreds, demonstrating just how poorly we capture extremely shadowed regions. The shadowed solar panels and facets sticking out from dimly lit portions of the spacecraft stand to gain the most from two-channel training, as optical observations reveal too little information about these characteristics for the final NeRF model to learn them.

The Hubble model is undoubtedly the most complex of the two geometries explored in this paper, but its visual performance was among the best observed. Figure 6 evaluates the full range trained model at its closest and furthest distances of 50m and 1km, respectively.



**(a) Evaluation at 50m** **(b) Evaluation at 1km**

**Figure 6. Interpolated PSNR, SSIM, and TIPE at 50m and 1km for the full-range Hubble model.**

Overall performance on the full range Hubble dataset was smooth throughout Figures 6 and 7—illuminated portions of the spacecraft performed very well, while the dim side showed the poorest performance. The final model outputs for three distinct lighting conditions are displayed in Figure 7. Across all lighting scenarios, MW-NeRF maintained excellent visible-band performance, although it did fail to capture a complete underlying shape model of the spacecraft in the shadowed region. In this case, the visible side is learned almost perfectly, while the shadowed side is completely absent (including the dark-side solar panel, apart from the portion that still reflects a bit of visible light).

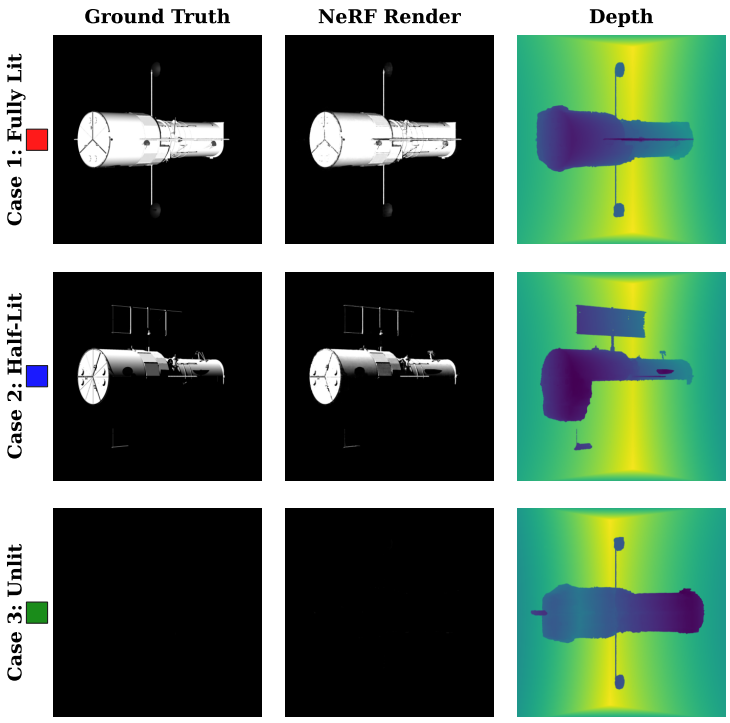



**Figure 7. Side-by-side comparisons of the learned Hubble model across the three lighting conditions labeled in the polar metric plots.**

### Initial IR Results

The current two-channel training with MW-NeRF has demonstrated great success in reconstructing accurate shape models while retaining good visible-band performance. This is best demonstrated in Figures 9 and 8.

Figure 9 shows our current visible-band accuracy while simultaneously learning an accurate underlying shape model. Comparing directly to Figure 6, we can see that supplementing IR data in the training process helps enhances model performance in some areas, while getting slightly worse in others in this case. While visible-only training fails to learn a large portion of Hubble, ultimately damaging the model's performance in that region, MW-NeRF can learn an accurate shape model fully, but currently still struggles with scene reconstruction in heavily shadowed zones. Figure 8 shows MW-NeRF's success in capturing the full Hubble shape model, especially when compared to case 2 in Figure 7. A full direct comparison between these two models when evaluated at a target range of 50m is shown in Table 2. As in the polar plots, we see extremely similar results across all metrics, with the MW-NeRF performance being slightly worse on average. The same inflated TIPE values in shadowed regions are shared between these two models, and while median PSNR and SSIM values remain within 2% of each other regardless of the hemisphere, we see median TIPE values differing by about 25-50%. Our current difficulties in the shadowed regions can be seen in case 3, where outlines of the spacecraft are learned to be slightly visible where they should not be. Overall, PSNR and SSIM results remained largely the same, while TIPE improved in some areas, and grew worse in the worst shadowed zones.

The model used to generate these results was trained on the same Hubble images as in Figure 6—

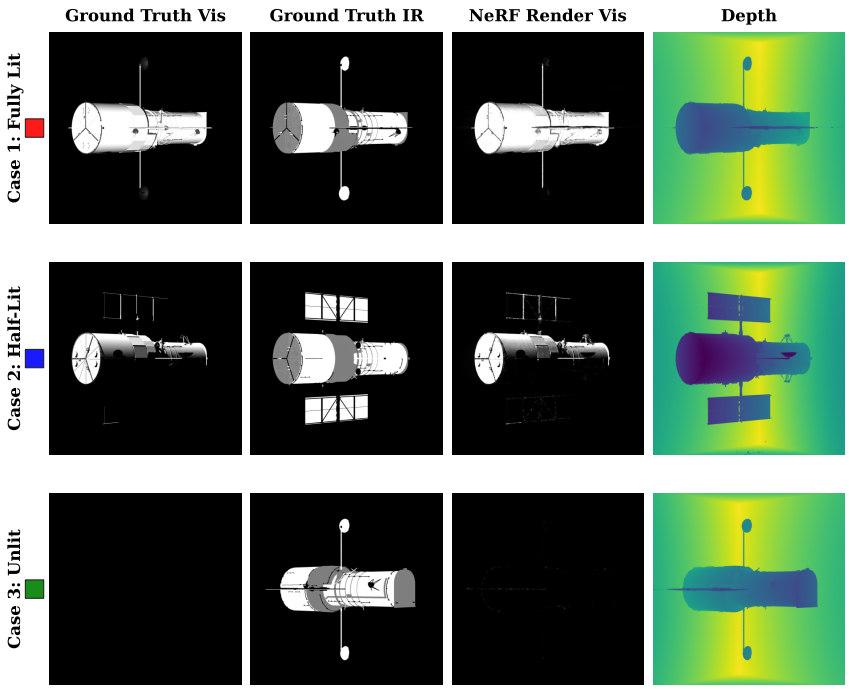
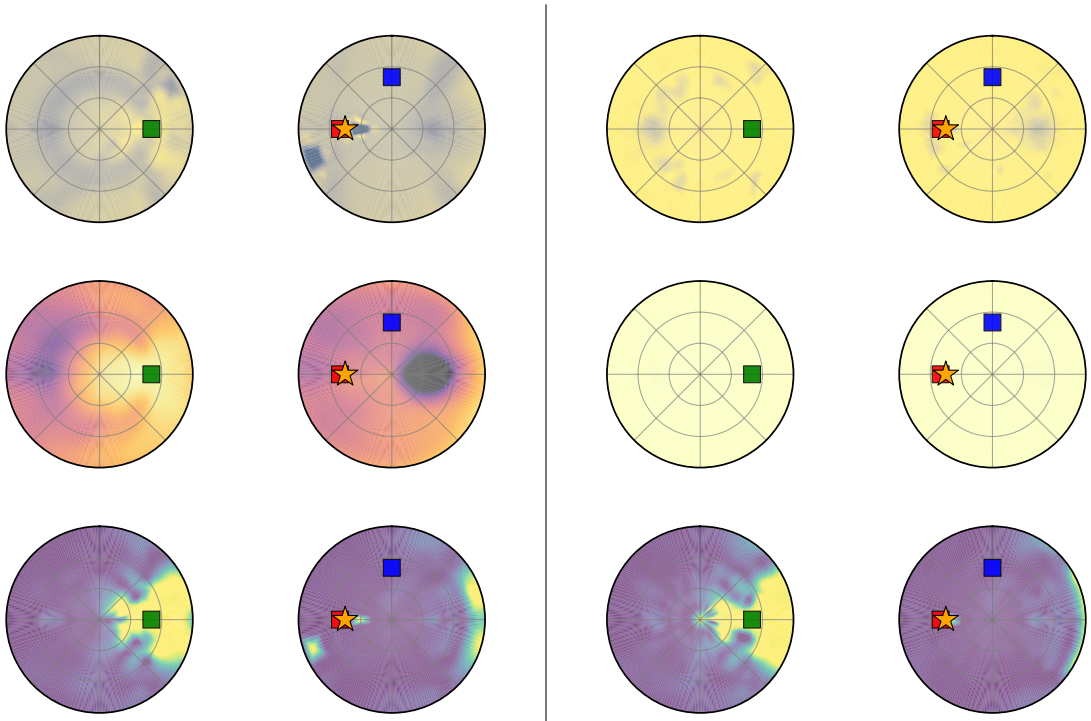



**Figure 8. Side-by-side comparisons of the learned Hubble model, using both vis and IR data, across three lighting conditions. Each NeRF render is accompanied by its corresponding depth map, showing the underlying learned structure of the scene.**

**Table 2. Hubble 50m Hemisphere Statistics for Vis-Only NeRF and MW-NeRF**

| Model | Metric | Hemisphere | Median | Min | Max | Recovered Shape |
|---|---|---|---|---|---|---|
| **Vis-Only NeRF** | TIPE | North | 4.4723 | 0.0091 | 713.489 | ✗ |
| | | South | 1.5091 | 0.0025 | 25.8616 | |
| | PSNR | North | 49.1113 | 26.0777 | 79.9753 | |
| | | South | 47.7291 | 22.9058 | 54.1801 | |
| | SSIM | North | 0.995997 | 0.9875 | 0.99996 | |
| | | South | 0.992505 | 0.9706 | 0.9969 | |
| **MW-NeRF** | TIPE | North | 6.0988 | 0.0117 | 761.8221 | ✓ |
| | | South | 2.5976 | 0.0252 | 479.8126 | |
| | PSNR | North | 48.8114 | 25.6463 | 66.6987 | |
| | | South | 47.2212 | 22.9883 | 53.5735 | |
| | SSIM | North | 0.994056 | 0.9830 | 0.99933 | |
| | | South | 0.990034 | 0.9676 | 0.9960 | |

using the same sun angle with target ranges from 50m to 1km. These 3070 visible-band images were accompanied by 614 IR images taken at the same grid points on the bounding sphere at a range of 50m. The loss associated with both visible and IR image renderings were weighted the same for this model's training, but further investigation is required to uncover what weights are appropriate based on the object geometry and lighting complexity. By exploring this, we expect to find weighting conditions which will maximize both shape model accuracy and visible-band performance, without compromising one or the other.



**(a) Evaluation at 50m** **(b) Evaluation at 1km**

**Figure 9. Interpolated PSNR, SSIM, and TIPE at 50m and 1km for the full-range Hubble model trained on visible and IR data.**

## CONCLUSION

In this work, we have outlined our new realistic datasets, which focus on two canonical geometries: a CubeSat model and a Hubble model. We also outlined and demonstrated the single-channel capabilities of our newly proposed NeRF model architecture: MW-NeRF. Our single-channel performance has demonstrated excellent capability in complex scenes with complex lighting, while still highlighting the fact that the harsh lighting conditions present in space-based observations are tremendously difficult to overcome using optical light alone. Early exploration of our two-channel capabilities has demonstrated that employing multi-wavelength data in NeRF training can significantly improve the learned shape model, but more work is needed to ensure this does not negatively affect visual performance. Future work will fine tune MW-NeRF when training on two channels to faithfully reconstruct accurate underlying shape models, while simultaneously producing more stable models that are globally accurate. Finally, producing accurate and reliable shape models for spacecraft paves the way for future work, which may require complete shape models, such as with dynamic relighting of scenes, or fully capturing time-varying scenes.

## ACKNOWLEDGMENT

We thank our collaborators from Johns Hopkins University Applied Physics Laboratory for helping to supply the training data used in our studies. We declare no conflict of interest.

## REFERENCES


[1] D. L. Oltrogge and S. Alfano, "The technical challenges of better space situational awareness and space traffic management," *Journal of Space Safety Engineering*, Vol. 6, No. 2, 2019, pp. 72–79.

[2] J. Lyzhoft, J. Swenson, D. Baker, A. Liounis, and M. Fickett, "On-Board Limb-Based Shape Modeling For Small Body Navigation," *3rd Space Imaging Workshop*, 2022.

[3] M. Ferrara, G. Arnold, and M. Stuff, "Shape and motion reconstruction from 3D-to-1D orthographically projected data via object-image relations," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 31, No. 10, 2008, pp. 1906–1912.

[4] P. Kou, Y. Liu, W. Zhong, B. Tian, W. Wu, and C. Zhang, "Axial Attitude Estimation of Spacecraft in Orbit Based on ISAR Image Sequence," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, 2021, pp. 7246–7258. Conference Name: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10.1109/JSTARS.2021.3096859.

[5] F. Wang, F. Xu, and Y.-Q. Jin, "3-D information of a space target retrieved from a sequence of high-resolution 2-D ISAR images," *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2016, pp. 5000–5002. ISSN: 2153-7003, 10.1109/IGARSS.2016.7730304.

[6] Q. Xu, M. Hu, Y. Fang, and X. Zhang, "A neural radiance fields method for 3D reconstruction of space target," *Advances in Space Research*, Feb. 2025, 10.1016/j.asr.2025.01.065.

[7] T. H. Park and S. D'Amico, "Rapid Abstraction of Spacecraft 3D Structure from Single 2D Image," *AIAA SCITECH 2024 Forum*, AIAA SciTech Forum, American Institute of Aeronautics and Astronautics, Jan. 2024, 10.2514/6.2024-2768.

[8] S. Lemmens, H. Krag, J. Rosebrock, and I. Carnelli, *RADAR MAPPINGS FOR ATTITUDE ANALYSIS OF OBJECTS IN ORBIT*. Apr. 2013.

[9] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," Aug. 2020. arXiv:2003.08934 [cs], 10.48550/arXiv.2003.08934.

[10] A. Mittal, "Neural Radiance Fields: Past, Present, and Future. arXiv. Cornell University," *arXiv preprint arXiv:2304.10050*, 2023.

[11] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4104–4113.

[12] R. Gaskell, O. Barnouin, M. Daly, E. Palmer, J. Weirich, C. Ernst, R. Daly, and D. Lauretta, "Stereophotoclinometry on the OSIRIS-REx mission: mathematics and methods," *The Planetary Science Journal*, Vol. 4, No. 4, 2023, p. 63.

[13] E. E. Palmer, R. Gaskell, M. G. Daly, O. S. Barnouin, C. D. Adam, and D. S. Lauretta, "Practical stereophotoclinometry for modeling shape and topography on planetary missions," *The Planetary Science Journal*, Vol. 3, No. 5, 2022, p. 102.

[14] S. J. Garbin, M. Kowalski, M. Johnson, J. Shotton, and J. Valentin, "FastNeRF: High-Fidelity Neural Rendering at 200FPS," Apr. 2021. arXiv:2103.10380 [cs], 10.48550/arXiv.2103.10380.

[15] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics*, Vol. 41, July 2022, pp. 1–15, 10.1145/3528223.3530127.

[16] R. Martin-Brualla, N. Radwan, M. S. M. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, "NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 7206–7215. ISSN: 2575-7075, 10.1109/CVPR46437.2021.00713.

[17] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields," Aug. 2021. arXiv:2103.13415 [cs], 10.48550/arXiv.2103.13415.

[18] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. Srinivasan, and J. T. Barron, "NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images," Nov. 2021. arXiv:2111.13679 [cs], 10.48550/arXiv.2111.13679.

[19] L. Yen-Chen, P. Florence, J. T. Barron, A. Rodriguez, P. Isola, and T.-Y. Lin, "INeRF: Inverting Neural Radiance Fields for Pose Estimation," 2021.

[20] C. Reiser, S. Peng, Y. Liao, and A. Geiger, "KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs," 2021.

[21] H. Turki, D. Ramanan, and M. Satyanarayanan, "Mega-NeRF: Scalable Construction of Large-Scale NeRFs for Virtual Fly- Throughs," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, IEEE, June 2022, pp. 12912–12921, 10.1109/CVPR52688.2022.01258.

[22] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields," Mar. 2022. arXiv:2111.12077 [cs], 10.48550/arXiv.2111.12077.

[23] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields," *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, IEEE, Oct. 2023, pp. 19640–19648, 10.1109/ICCV51070.2023.01804.

[24] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "D-NeRF: Neural Radiance Fields for Dynamic Scenes," 2020.

[25] S. Fridovich-Keil, G. Meanti, F. R. Warburg, B. Recht, and A. Kanazawa, "K-Planes: Explicit Radiance Fields in Space, Time, and Appearance," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, IEEE, June 2023, pp. 12479–12488, 10.1109/CVPR52729.2023.01201.

[26] S. Chen, B. Wu, H. Li, Z. Li, and Y. Liu, "Asteroid-NeRF: A deep-learning method for 3D surface reconstruction of asteroids," *Astronomy & Astrophysics*, Vol. 687, July 2024, p. A278. Publisher: EDP Sciences, 10.1051/0004-6361/202450053.

[27] B. Caruso, T. Mahendrakar, V. M. Nguyen, R. T. White, and T. Steffen, "3D Reconstruction of Non-cooperative Resident Space Objects using Instant NGP-accelerated NeRF and D-NeRF," June 2023. arXiv:2301.09060 [cs], 10.48550/arXiv.2301.09060.

[28] A. Mergy, G. Lecuyer, D. Derksen, and D. Izzo, "Vision-based Neural Scene Representations for Spacecraft," May 2021. arXiv:2105.06405 [cs], 10.48550/arXiv.2105.06405.

[29] A. M. Heintz and M. Peck, "Spacecraft State Estimation Using Neural Radiance Fields," *Journal of Guidance, Control, and Dynamics*, Vol. 46, No. 8, 2023, pp. 1596–1609. Publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/1.G006946, 10.2514/1.G006946.

[30] A. Legrand, R. Detry, and C. D. Vleeschouwer, "Leveraging Neural Radiance Fields for Pose Estimation of an Unknown Space Object during Proximity Operations," June 2024. arXiv:2405.12728 [cs], 10.48550/arXiv.2405.12728.

[31] G. B. Palmerini, "Combining thermal and visual imaging in spacecraft proximity operations," *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, IEEE, 2014, pp. 383–388.

[32] Y. Y. Lin, X.-Y. Pan, S. Fridovich-Keil, and G. Wetzstein, "ThermalNeRF: Thermal Radiance Fields," *2024 IEEE International Conference on Computational Photography (ICCP)*, IEEE, 2024, pp. 1–12.

[33] K. Kurata, H. Niigaki, X. Wu, and R. Tanida, "MultiBARF: Integrating Imagery of Different Wavelength Regions by Using Neural Radiance Fields," Mar. 2025. arXiv:2503.15070 [cs], 10.48550/arXiv.2503.15070.

[34] C.-H. Lin, W.-C. Ma, A. Torralba, and S. Lucey, "Barf: Bundle-adjusting neural radiance fields," *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 5741–5751.